

Technical Note: A Functional View of Context-Driven Learning

Core Idea

A context-driven learning system can be viewed as the interaction of three coupled components:

- **Context generation**
- $C(x)$
- **Conditional computation**
- $P(C, x)$
- **Learning**
- L

Rather than viewing a neural network as a single nonlinear function, this viewpoint separates the architecture into three interacting processes.

1. Context Generation

The context function

$$C(x)$$

maps an input into a computational state.

The context may consist of:

- ReLU activation patterns,
- binary context vectors,
- LSH signatures,
- random projection codes,
- decision paths,
- routing variables,
- or any other representation describing the current computational situation.

The context answers one question:

Where in computational state space is this input?

2. Conditional Computation

The conditional computation operator

$$P(C, x)$$

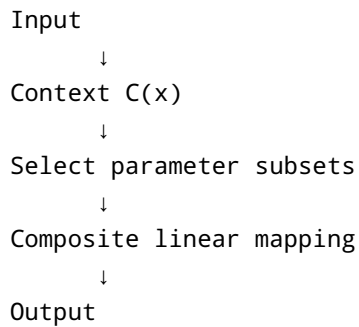
computes the output using both the input and its computational context.

In the simplest case,

$P(C, x)$ is a composite linear mapping.

The context selects which parameter blocks participate in the computation, producing an effective linear operator for the current input.

Conceptually,



Each input therefore experiences a different effective computation.

The network behaves as though it contains a very large collection of local linear predictors that share parameters extensively.

The selected mapping is not stored explicitly.

It is assembled dynamically from shared parameter blocks according to the current context.

3. Learning

Learning is performed by

$$L.$$

Backpropagation repeatedly observes errors produced by the current conditional computation.

These errors are propagated through the selected computational pathway and used to update the shared parameters.

Over many training examples, backpropagation gradually discovers parameter values that work well across many different contexts.

Rather than optimizing one fixed linear mapping, learning organizes an enormous family of context-dependent mappings that share computational structure.

Computational Factorization

The complete task may be far too complicated to learn as one global computation.

Instead,

- each input selects its own composite computation,
- related inputs share many parameter blocks,
- different contexts reuse overlapping computational components,
- backpropagation coordinates all parameter updates simultaneously.

The overall computation therefore becomes factorized into reusable pieces.

This factorization is not specified manually.

It emerges through learning.

Why Parameter Sharing Works

Parameter sharing is possible because similar contexts often require similar computations.

Backpropagation gradually identifies these similarities.

Parameters that prove useful across many contexts are strengthened.

Parameters useful only for specialized situations naturally become specialized.

Learning therefore balances:

- reuse,
 - specialization,
 - generalization.
-

Beyond Linear Predictors

The composite linear mapping provides the simplest mental model.

However,

$$P(C, x)$$

need not be linear.

The conditional computation could instead be:

- another neural network,
- a transformer block,
- a mixture-of-experts module,
- a dynamical system,
- a probabilistic model,
- or any context-dependent computational operator.

The essential idea remains unchanged.

The context determines which computation is performed.

Learning organizes that computation through repeated error correction.

Summary

This functional decomposition separates three fundamentally different roles within a learning system.

- $C(x)$ constructs computational coordinates.
- $P(C, x)$ performs computation conditioned on those coordinates.
- L organizes the shared parameters through backpropagation so that the conditional computations collectively approximate the desired mapping.

From this viewpoint, learning is not simply the optimization of a nonlinear function. It is the continual organization of a large family of context-dependent computations into a compact, factorized system of shared parameters.